

# Free speech in the age of online content moderation

---

Giovanni de Gregorio

2019-11-26T10:00:34

Content moderation is not a novelty in the media sector. As content providers, traditional media outlets like televisions and newspaper have always selected the information to broadcast or disclose. Nevertheless, when social media companies are involved in this activity, content moderation raises serious [concerns](#). In the last years, the digital environment has been playing a crucial role in promoting the sharing of opinion and ideas on a global scale. Nevertheless, this flourishing democratic framework firmly clashes with the troubling evolution of the algorithmic society where social media platforms govern the flow of information. It would be enough to observe that, just in the case of Facebook or YouTube, the amount of posts moderated in different areas of the world is [on a scale of billions](#) each week. This situation explains why content moderation is usually performed by machines which can efficiently decide in a heartbeat whether to maintain or delete the vast amount of content flowing every day on social media.

## Free speech under pressure

In this scenario, it is not by chance that the activity of content moderation performed by social media companies on a global scale challenges the protection of free speech in a twofold way. Firstly, the activity of content moderation affects freedom of expression since social media companies can select which information deserves to be maintained and deleted according to standards based on the interest to avoid any monetary penalty or reputational damage (ia collateral censorship). Moreover, an extensive activity of content moderation influences even the right to privacy and data protection. Indeed, users could fear to be subject to a regime of private surveillance of their information and data, thus adapting their behaviours in light of these threats. Secondly, content moderation has shown to affect indirectly other fundamental rights or, from an international law perspective, human rights. It would be enough to focus on the role of social media companies in escalating violent conflicts in countries like [Myanmar](#) or [Sri Lanka](#) where genocide and mass atrocities occurred. Some states, increasingly in [Africa](#), even decided to shutdown social media.

## Whom to address?

This troubling situation is the result of the discretion social media companies enjoy in deciding how to moderate content by interpreting users' right to free speech according to their own legal, economic and ethical framework. Social media companies have traditionally been considered as passive service providers enjoying an exemption of liability for unlawful content uploaded by third parties. However, as I have already tried to explain in one of my [articles](#), this legal picture is far from reality, where online platforms started to exercise new forms of powers. To some extent, the result of this privately-driven activity mirrors the exercise of judicial balancing and

public enforcement carried out by state actors. Even more importantly, the activity of content moderation contributes to privately shaping the boundaries of the right to free speech online, overcoming the principle of the rule of law by proposing a private standard of protection. For instance, the removal or blocking of online content is enforced directly by social media companies, who are deciding which right should prevail in each specific case based on their own, internal criteria. The situation is still more compelling when considering that social media companies outsource content moderation and this activity is performed by employees who – around the world and without legal training – do not spend more than a few seconds for each content or, even less, when machines are involved.

Nevertheless, social media companies, as private actors, are not obliged to respect fundamental rights since their protection can be enforced only vis-à-vis states in the lack of any regulation of content moderation. Although these actors are usually neither accountable nor responsible for hosting third-party content, nevertheless, their decisions can shape the right to free speech, especially when assessing users' requests to remove flagged content, which illicit nature is not always evident – for example, if it concerns disinformation. Despite their crucial role in the digital environment, social media companies do not ensure transparency and explanation of their decision-making processes. Users cannot rely on any rights in regard to how social media companies moderate their content. Indeed, users cannot access the reasons why a specific content has been removed or ask social media for a review of their previous decision.

### **How to address the issue?**

Within this framework, the regulation of users' rights in the field of online content moderation could be the first step to address the threats for fundamental rights and mitigate platforms' power. As I have tried to [explain](#), this approach would be based on a regulation of the process which avoids the imposition of monitoring obligations and rather focuses on transparency and accountability safeguards in online content moderation. Indeed, in order to avoid interferences with fundamental rights, especially the right to free speech, the regulation of content moderation should not lead to an obligation of online platforms to generally monitor online content. Moreover, content moderation procedures should be explained to users in a transparent and user-friendly way. In these cases, the principle of human-in-the-loop in content moderation plays a crucial role, since humans could be an additional safeguard for users to rely on a human translation of the procedure. However, this regulatory strategy should strike a fair balance between the rights of the users and the freedom to conduct business of online platforms. The principle of proportionality in online content moderation should guide any attempt of regulation in this field. Indeed, such new digital rights should not be absolute but subject to conditions not substantially impeding social media companies to run their business. This is also why, as addressed by the CJEU in [Google v. CNIL](#), the standard of the right to free speech is not the same worldwide, thus, requiring, to thoroughly take into account in the context of global removal (under the conditions laid out in [Glawischnig-Piesczek v. Facebook](#)). Despite these differences at the international level, in any case, new users' rights in the field of online content would be a crucial

step to ensure that social media companies, as private actors, do not impose their standard of fundamental rights protection in the digital environment.

*Giovanni de Gregorio is an Academic Fellow and PhD Candidate in Public Law at the Department of Law at the Bocconi University at Milan.*

Cite as: Giovanni de Gregorio, “Free Speech in the Age of Online Content Moderation”, *Völkerrechtsblog*, 26 November 2019.

